

TAIGA: A Novel Dataset for Multitask Learning of Continuous and Categorical Forest Variables From Hyperspectral Imagery

Matti Möttöus¹, Phu Pham, Eelis Halme, Matthieu Molinier², *Member, IEEE*, Hai Cu, and Jorma Laaksonen, *Senior Member, IEEE*

Abstract—The spectral and spatial resolutions of modern optical Earth observation data are continuously increasing. To fully utilize the data, integrate them with other information sources, and create applications relevant to real-world problems, extensive training data are required. We present TAIGA, an open dataset including continuous and categorical forestry data, accompanied by airborne hyperspectral imagery with a pixel size of 0.7 m. The dataset contains over 70 million labeled pixels belonging to more than 600 forest stands. To establish a baseline on TAIGA dataset for multitask learning, we trained and validated a convolutional neural network to simultaneously retrieve 13 forest variables. Due to the size of the imagery, the training and testing sets were independent, with strictly no overlap for patches up to 45 × 45 pixels. Our retrieval results show that including both spectral and textural information improves the accuracy of mapping key boreal forest structural characteristics, compared with an earlier study including only spectral information from the same image. TAIGA responds to the increased availability of hyperspectral and very high resolution imagery, and includes the forestry variables relevant for forestry and environmental applications. We propose the dataset as a new benchmark for spatial–spectral methods that overcomes the limitations of widely used small-scale hyperspectral datasets.

Index Terms—Boreal forest, convolutional neural networks, hyperspectral imaging (HSI), multitask learning.

I. INTRODUCTION

SATELLITE-BASED remote sensing is a widely accepted tool in forest inventory and for monitoring the extensive forests covering approximately one-third of the world’s land surface. This essential service is largely based on high- to medium-resolution multispectral optical instruments with pixel sizes of 10 m or more. While the availability of imagery with better spatial resolution, 1 m or less [called very high resolution (VHR)] is increasing rapidly, its utility in forestry

applications has not been demonstrated yet [1]. VHR images are useful in visual analyses, but lack the large-scale automated processing chains required for routine forest applications.

A natural approach to automate VHR image processing is deep learning, building on its breakthroughs in computer vision and speech recognition. Over the past few years, it has been increasingly applied in Earth observation (EO) [2]. Deep learning methods have been successfully adapted to the spectral, spatial, and temporal specificities of EO imagery [3], [4]. In parallel with the ongoing enhancements in spatial image resolution, hyperspectral imaging (HSI) aims for the same in the spectral dimension, collecting continuous reflectance spectra using hundreds of narrow reflectance bands, providing the best discriminating power in the spectral domain [5]. The imagery produced by a HSI sensor is often viewed as a data cube with three continuous axes: one spectral and two spatial. Building upon the legacy of earlier machine learning-based spatial–spectral methods [6]–[8], deep learning has also been applied to HSI data cubes, with convolutional neural networks (CNNs) being the most popular approach [9]. CNN models can extract discriminatory features from the data cube and exploit its spectral and spatial information. The use of both spectral and spatial information for hyperspectral image analysis has been shown to improve the accuracy in image classification [10]–[12]. Several reviews are already available on machine learning in HSI analysis [13]–[15], indicating that the field has already an established status.

Hyperspectral image classification is the most popular method for scene analysis from HSI [16]. Here again, CNNs have been found to be more accurate compared with other deep learning methods or the more traditional classification algorithms, e.g., support vector machines (SVMs) [10], [17]. Recently, 3-D CNN architectures, which use both spectral and spatial information, have been used for the classification of hyperspectral data [12], [18]. In addition, novel deep CNNs that integrate both spatial context and spectral signature have been designed for HSI classification in a few studies [11], [19]. Despite numerous works on deep learning for HSI, there remain research gaps and a fundamental limitation related to commonly used HSI datasets.

Prediction of continuous biophysical variables from HSI EO data with CNNs is not yet well studied [20], [21]. Still, these data–method combination has been used in a few studies. For instance, Liu *et al.* [22] achieved good accuracy in predicting

Manuscript received April 13, 2021; revised September 10, 2021 and December 2, 2021; accepted December 26, 2021. Date of publication January 7, 2022; date of current version March 11, 2022. This work was supported in part by the Academy of Finland AIPSE Research Program (AIROBEST Project) under Grant 317387 and in part by the Academy of Finland General Funding under Grant 322256. (*Corresponding author: Matti Möttöus.*)

Matti Möttöus, Eelis Halme, and Matthieu Molinier are with the VTT Technical Research Centre of Finland Ltd., 02044 Espoo, Finland (e-mail: matti.mottus@vtt.fi).

Phu Pham was with the Department of Computer Science, Aalto University, 00076 Aalto, Finland. He is now with Purdue University, West Lafayette, IN 47907 USA.

Hai Cu and Jorma Laaksonen are with the Department of Computer Science, Aalto University, 00076 Aalto, Finland.

Digital Object Identifier 10.1109/TGRS.2022.3141217

soil clay content with a 1-D CNN approach (the coefficient of determination, $R^2 = 0.83$). CNN has also been found to have great potential for cyanobacteria detection and quantification with HSI [23].

Multitask learning, defined as jointly solving a set of prediction problems by sharing information across tasks [24], has attracted even less attention in scientific publications, but the situation is improving rapidly [25], [26]. In practical EO, it corresponds to the joint classification and regression (variable estimation) with a training dataset of mixed continuous and categorical variables. Furthermore, we chose multitask learning because in forest mapping and many other real-world problems, a coherent retrieval of not only target classes, but also numerous parameters for each detected class, is essential. With a multitask learning approach, the model has an opportunity to learn or take into account those hidden correlations between variables. Multitask learning also saves time and computing resources compared with training a separate CNN model for each variable.

In HSI analysis, the viability of multitask learning has been demonstrated in superpixel-based classification [27] and combining samples from different hyperspectral images to improve classification accuracy [25]. However, the machine learning methods used in EO usually only make use of the spectral information [28] with multitask learning used to improve image classification. The full potential of multitask machine learning in EO, and HSI analysis in general, is yet to be unleashed.

Whereas numerous machine learning and deep learning methods have been proposed to make use of the spatial and spectral characteristics of hyperspectral imagery, the key limitations of the approach are the adequate training of the model and a statistically sound accuracy assessment on an independent dataset. The lack of large-scale labeled HSI datasets has pushed practitioners to rely heavily on a limited selection of small aerial imagery (e.g., Indian Pines and Salinas: University of Pavia [25], [27]), each consisting of a single image with only a few hundred pixels in each spatial dimension. Training and evaluating models on such small images with randomly sampled large patches as inputs leads to spatial overlap between training and testing sets (directly or through the receptive field of deep networks), resulting in a lack of independence between training and testing sets, and ultimately to over-optimistic accuracy assessment [14]. Although alternative sampling schemes have been proposed to mitigate these effects [29]–[32], these cannot fully overcome the physical limitation caused by the small number of pixels. Fixed data splits between training and testing sets have been introduced for these datasets by IEEE Geoscience and Remote Sensing Society (GRSS)¹, which further reduce the amount of data for training models and do not completely solve the overlap when larger patches are used. Larger HSI datasets have also been released, such as the Houston University datasets from GRSS Data Fusion Contests in 2013 [33] and 2018 [34], and the Chikusei dataset [35]. These datasets include fixed

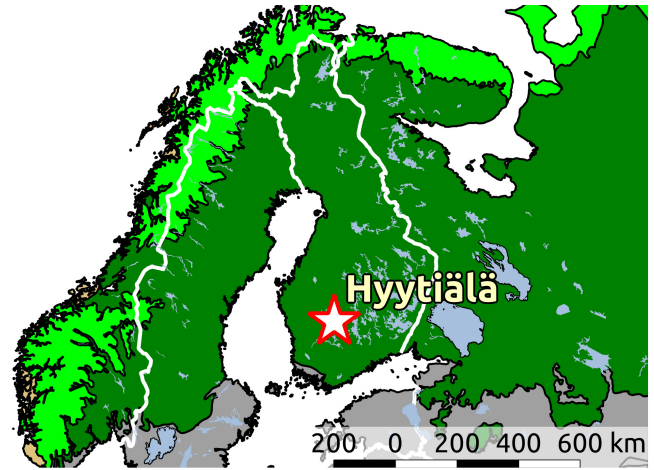


Fig. 1. Location of the test site in European boreal forest zone (dark green color).

disjoint training and testing sets that mitigate the overlap issue, yet many pixels are left unlabeled.

The motivation of the manuscript is the need for a large-scale HSI dataset, which would enable both deep-learning based multitask learning and a statistically sound accuracy assessment on an independent test set. We present TAIGA, an openly available VHR HSI dataset covering the visible to near-infrared spectral regions (VNIR, 400–900 nm) acquired above a boreal forest in Central Finland. The imagery, offering 70 million labeled pixels, is accompanied by extensive and continuous stand-wise forestry data from regularly updated nationwide measurements for over 600 forest stands. It exceeds the current standard benchmark hyperspectral datasets both in image size and the amount of available labeled data, and enables the joint classification and regression of forest variables. We obtain baseline results for this unique dataset by applying CNNs for multitask learning of the forest variables using strictly independent samples for testing and training. In addition, we compare the results with a previous exercise on forest variable retrieval using only spectral information, therefore quantifying the advantages of combining spectral and spatial features in the analysis of VHR HSI data in forestry applications.

II. DATA

A. Study Site

The HSI data were collected in the southern boreal zone of Finland, in the vicinity of Hyytiälä forestry field station (61°50'44"N, 24°17'10"E, Figs. 1 and 2). The study area is mostly covered by managed boreal forests, agricultural fields, and wetlands. The main overstory species of the forests are, in the order of dominance, Scots pine (*Pinus sylvestris*), Norway spruce (*Picea abies*), and Silver birch (*Betula pendula*), either as mixed or single-species stands growing on mostly Podzolic soils. As a result of management, forest density varies usually between 500 and 1500 trees/ha. Different shrubs,

¹IEEE Data and Algorithm Evaluation Website (DASE) - <http://dase.grss-ieee.org>



Fig. 2. Aerial view of the landscapes around the Hyytiälä forestry field station.

lichens, and mosses cover the forest floor, no bare soil is visible. The 30-year average annual precipitation at Hyytiälä is 711 mm, and the annual mean temperature is 3.5 °C [36]. The area has a mean elevation of approximately 150 m above sea level (asl) and does not exhibit large variations in elevation, with some undulating hills reaching less than 200 m asl.

B. Hyperspectral Data

Airborne data were collected with the AISA Eagle II hyperspectral line scanner [128 bands, 400–1000 nm, spectral resolution 4.7-nm full width half maximum (FWHM)] from a height of approximately 1 km above ground. The flight configuration, the field of view of the scanner (37.7°), and the number of across track pixels (1024) led to a spatial resolution of 0.7 m. Data were obtained from seven flight lines flown parallel to the direction of sunrays, under a clear sky on June 15, 2017. The solar zenith angle decreased from 42.3° to 40.3° during the data acquisition between 11:37 A.M. and 12:14 P.M. (Eastern European Summer Time, UTC+3). The hyperspectral data were projected to UTM geographical coordinates using satellite and inertial positioning data and a digital elevation model based on nationwide laser scanning data with the Atcor software package (ReSe Applications GmbH, Switzerland) and resampled onto an orthogonal grid. The data were further converted to top-of-canopy reflectance with the Atcor package (ReSe Applications GmbH, Switzerland) using sun photometer-measured aerosol optical properties provided by the the AEROSOL ROBOTIC NETWORK (AERONET, <https://aeronet.gsfc.nasa.gov>) and mosaicked. A comprehensive description of the data-preprocessing chain applied to the airborne data is reported by Markiet *et al.* [37]. The measured area was approximately 9 km by 3 km. Based on a comparison with fixed ground points, the geolocation accuracy of the data was approximately 2–3 m, determined mostly by the accuracy of the top-of-canopy digital surface model and the inertial measurement system used during the acquisitions.

C. Finnish Forestry Data

We used field data based on the forest resource information gathered and processed by the Finnish Forest Center (FFC) on privately owned Finnish forests. The inventory process is executed by FFC as a statutory assignment together with a few Finnish forestry and remote sensing companies. General outline of the laser scanning based stand level inventory is described by Maltamo *et al.* [38] and a more detailed description of the inventory process by Holopainen *et al.* [39]. The inventory process is summarized below.

The forest data are collected by nationwide field measurements and airborne remote sensing—Airborne Laser Scanning (ALS) and aerial imaging. The number of reference plot measurements per one inventory area, approximately 100 000–200 000 ha, is about 600–800. The plots typically have a radius of 9 m, and variables measured in the field for each species are mean height, basal area, stem count, mean diameter, and age. The reference plots are distributed to represent the existing forest variation well. The ALS campaign is performed during a single growing season (leaf-off or leaf-on). If the ALS data are acquired on a different year than the field measurements, the reference plots can be grown with models to match the correct epoch. The aerial imagery are four-channel images (visible and near-infrared).

In the inventory process, the numerical forest variables are imputed from ALS data for an inventory unit, a 16 m × 16 m cell. The units form a nationwide grid. Statistical nonparametric models are built for each variable for predicting them from the height distribution and the intensity of the ALS data, calibrated with the field plot data. The texture metrics and the spectral statistics of the aerial imagery are used and auxiliary data source, e.g., in species determination. Due to their empirical nature, models are specific to inventory areas. In addition, information about different forestry related soil properties (e.g., soil type and fertility class) are retrieved for the grid cells from external data sources.

The stand-wise data by FFC contains information characteristic to the stand (e.g., soil type, fertility class) or to a class of trees growing inside the stand, called a “stratum.” Each stratum quantifies the contribution of trees similar in species, age, height, etc., to the forest stand. The cumulative characteristics of a stand, such as its basal area, above-ground woody biomass or stem volume, were obtained as sums of the corresponding values of the strata; mean tree height and diameter at breast height (DBH) were computed as basal-area weighted average of the tree heights of all strata in a stand. Similarly, the species composition of a stand was calculated on the basis of basal area. The forest stand boundaries were contracted by 10-m buffer in further processing to avoid geolocation errors and border or neighborhood effects in the hyperspectral data.

In Finnish forestry, forests and their resource data are managed on stand level. A stand is an aggregation of trees that are sufficiently uniform in species composition, size, arrangement, and age. Definition of stands starts by a segmentation of forest area into “micro stands” using canopy height model and aerial imagery. Forest stands are then delineated by combining

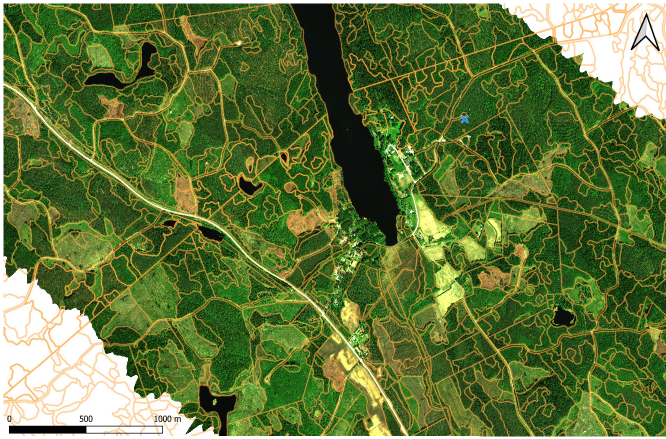


Fig. 3. RGB image of a subset of the mosaicked hyperspectral data, overlaid by the stand boundaries of the region (orange).

similar neighboring micro stands into economically viable forest units (Fig. 3).

The stand-level forest resource data are aggregated from the grid cells using the vector files with stand geometries. If one stand geometry includes enough complete or nearly complete grid cells, the variables of the stand are computed using these grid cells. If not enough grid cells fall inside the stand, some partly intersecting grid cells are also taken into account using their weighted averages, decreasing the accuracy of the stand variables. The weight comprises two elements: grid cell surface area proportion within a stand and a weight factor describing the data reliability of a grid cell. The latter takes into account the poorer reliability of a cell that partly intersects the stand compared to one falling completely inside it.

Forests in different parts of the country are measured each year. The nation-wide forest variable data are continuously updated based on these measurements, forest management notices, and empirical forest growth models. We used the forestry data with a reference date at the end of year 2017. FCC provides the nationwide forest data as a free download with an open license (Creative Commons Attribution 4.0 International, CC BY 4.0) on a website maintained by FCC (<https://www.metsaan.fi/en/spatial-datasets>).

D. Artificial Intelligence Dataset for Forest Geographical Applications (TAIGA)

The classification of site fertility in Finnish forestry is based on forest floor vegetation [40]. A majority of the study sites belonged to the following four fertility classes, listed in decreasing order of fertility: herb-rich, mesic, sub-xeric, and xeric heath forests [41] (codes 2–5 in the Finnish system). The few stands belonging to other classes did not constitute a sufficient sample and were merged with their closest class among the four listed above—either herb-rich heath or xeric forest. Additionally, we significantly reduced the number of soil types given in the forestry database. In Finnish boreal forests, soil is covered by mosses, herbs, lichens, and shrubs and will not have a direct effect on the reflectance signal. We therefore merged the ten distinct soil types in the database

into two: mineral and organic soils (coded as 1 and 2 in the data file).

Using the FCC data, we calculated the percentages of pine, spruce, and broadleaf (i.e., birch) in each stand based on their basal areas. The species with the largest basal area was considered the main tree species (note that this differs slightly from the definition used by the FCC data). For each stand, we retrieved the key forest variables, mean diameter at breast height (1.3 m, DBH), stem density (trees per hectare), basal area (cross-section area of tree trunks at 1.3 m, in square meters per hectare), and mean tree height.

In addition to the traditional forest variables given in the database, we estimated the leaf area index (LAI) of each stand using the forestry data and the biometric regressions published by Repola [42], [43] for the Finnish boreal forest. We used the species-specific leaf area and shoot silhouette to total area ratios suggested by Majasalmi *et al.* [44] for the same geographic area, to obtain the total LAI and shoot-corrected effective LAI for each stratum. The effective LAI values take into account the effect of the clumping of conifer needles into shoots on the transmission of incident sunlight and were considered as potential predictors of the reflective properties of forests. For broadleaves, effective LAI was equal to the LAI. Note that the effective LAI values used here are abstract concepts and are not supposed to correspond to the optical canopy transmission measurements as only shoot-level clumping was considered; clumping of leaves around branches and into crowns was ignored due to a lack of reliable information. The list of the stand characteristics used in the study is given in Table I. The dataset containing the hyperspectral image, stand boundaries, and associated forest variables will hereafter be referred to as Artificial Intelligence dataset for Forest Geographical Applications (TAIGA). The Python code used to process the forest variables in TAIGA from the open data by FCC is available on GitHub in the *taiga-baseline* directory in <https://github.com/aalto-cbir/AIROBEST>.

Part of the stands within the hyperspectral image were not suitable for further analyses. We created a list of unsuitable stands using thresholds for minimum area, the normalized difference vegetation index (NDVI) computed from red- and near-infrared reflectance [45], and effective LAI. The LAI and NDVI thresholds, 0.86 and 0.61, respectively, were selected visually and the minimum area of 0.5 ha is a general recommendation from FCC. The LAI threshold excluded stands that were logged between the flight campaign (June 2017) and the epoch of the forest data (end of 2017); the NDVI threshold excluded additional recently logged stands with little leaf cover. The unsuitable stands were mapped as “no data” in TAIGA.

The TAIGA dataset (<https://doi.org/10.23729/fe7ce882-8125-44e7-b0cf-ae652d7ed0d5>), including the hyperspectral image, forest variable data, and metadata are available through the Fairdata IDA service (<https://etsin.fairdata.fi>), a continuous service for safe research data storage organized by the Ministry of Education and Culture of Finland. The data are stored as geolocated rasters in ENVI format. The data are freely available under Creative Commons Attribution 4.0 International (CC BY 4.0) license.

TABLE I

VARIABLES IN TAIGA. FFC REFERS TO DATA DIRECTLY PROVIDED BY THE FINNISH FOREST CENTER, AN ASTERISK DENOTES A DERIVED QUANTITY. FOR ALLOMETRIC RELATIONSHIPS, SEE THE TEXT

#	Variable name	No. of classes or range	Source
<i>Categorical</i>			
1	Fertility class	4	FFC
2	Soil class	2	FFC*
3	Main tree species	3	FFC
<i>Continuous</i>			
4	Basal area [m ² /ha]	0 – 35.51	FFC
5	Mean DBH [cm]	0 – 30.89	FFC
6	Stem density [1/ha]	0 – 6240	FFC
7	Mean height [m]	0 – 24.16	FFC
8	Percentage of pine	0% – 100%	FFC*
9	Percentage of spruce	0% – 84%	FFC*
10	Percentage of birch	0% – 58%	FFC*
11	Woody biomass [tn/ha]	0 – 180	FFC
12	Leaf area index	0 – 9.66	allometry
13	Effective leaf area index	0 – 6.45	allometry

TABLE II

NUMBER OF STANDS AND PIXELS IN THE TRAINING, VALIDATION, AND TEST SETS

set	stands	samples
training	> 560 <	37749
validation		4195
test	111	20263

III. METHODS

A. Data Preprocessing for the CNN

We predicted three categorical and ten continuous forest variables (Table I) from the first 110 bands in the airborne HSI data. The bands with wavelengths above 910 nm were discarded due to their high noise level.

To make the hyperspectral and forest data structurally similar, we created raster images of the forest variables, where the values of all forest variables were constant within the same stand, with the exception of 10-m buffers, in line with the processing of the hyperspectral imagery. The continuous variables were normalized between 0 and 1 after clipping to the 98th percentile.

The stands were split into training, validation, and test sets (Table II) so that the training and testing samples were never overlapping, and no single pixel appeared in both training and testing sets. The input image patch size used by the 3-D CNN was 45 × 45 pixels. The patch size was selected large enough to include relevant spatial context for the analysis of submetric hyperspectral images. We tried several sizes, from 27 × 27 to 91 × 91 pixels, with 45 × 45 offering the best compromise quantified by the accuracy measures (RMSE and rRMSE) for basal area and LAI, while still strictly forbidding any spatial overlap between training and test sets selected within the TAIGA dataset [46], [47]. The center pixels of the patches were picked from the stands with stride of 13 × 13 pixels so that a 27 × 27 image window centered in the sample center was entirely inside the buffered borders of a forest stand. We ignored all nonforest pixels (i.e., water, roads, agricultural fields, and residential areas).

B. TAIGA Hyperspectral Analysis CNN Pipeline

Our baseline 3-D CNN pipeline was inspired by the models studied by Chen *et al* [20], who found out that 3-D CNN could simultaneously utilize both spatial and spectral features from the hyperspectral image. For this work, we also studied and experimented with various other CNN models that were designed for hyperspectral imagery, including Lee’s model [19] and He’s model [48]. These models performed well on single-task learning on small datasets. However, the performance degraded significantly when applied to multi-task learning problems. Among these models, Chen’s model showed the most promising results, while the Lee’s and He’s models provided clearly inferior results.

The similarities between our model and Chen’s model include the use of the same feature extraction part (up to the shared FC layer in our architecture) that utilizes 3-D convolution operations, and the use of the same data augmentation methods. Considering the complex nature of the data and tasks, aside from those similarities, our model has some major differences in comparison to Chen’s model. Our model has separate output layers for multitask learning that can predict both discrete and continuous variables. Besides oversampling minority classes, we also applied different approaches to tackle the class-imbalanced problem, including cost-sensitive learning, class rectification loss [49], and focal loss [50]. Since our model was designed for multitask learning, we applied GradNorm [51] and uncertainty loss [52] to balance learning among specified tasks. We also carried out extensive numerical experiments to find the optimal choices for the 3-D CNN model configuration, including the image patch size, number of convolution kernels, kernel sizes, various regularizations, and sampling techniques. The detailed experiments and results of these experiments are presented in [53].

In this article, we describe the best-performing variant of our experiments as the TAIGA baseline method. The pipeline of the architectures is illustrated in Fig. 4. We adopted the hard parameter-sharing approach for our multitask learning problem, which means that the pipeline includes two parts: shared layers and task-specific layers.

The shared part of the pipeline consists of four convolutional (Conv) layers (Conv1 to Conv4 in Fig. 4) interleaved with two pooling (Pool) layers and followed by a shared fully connected (FC) layer. At each Conv layer, we used a stride of one and a padding of zero for the 3-D convolution operation. For the pooling operation, a kernel of size 2 × 2 × 1 was used. After each pooling layer, we applied the nonlinear ReLU activation function.

We used kernels with sizes of 3 × 3 × 54, 3 × 3 × 32 and 3 × 3 × 32 for the three simple convolution layers Conv1, Conv3, and Conv4, respectively, in our baseline model. We adopted the idea of multiscale convolutions in our model, and used a multiscale convolution block that comprises four convolution layers with the kernel sizes of 1 × 1 × 1, 1 × 1 × 3, 1 × 1 × 5, and 1 × 1 × 11, respectively. These convolution filters were used to exploit the spectral correlations. No pooling was applied after these layers. The outputs of the convolution layers were then summed together.

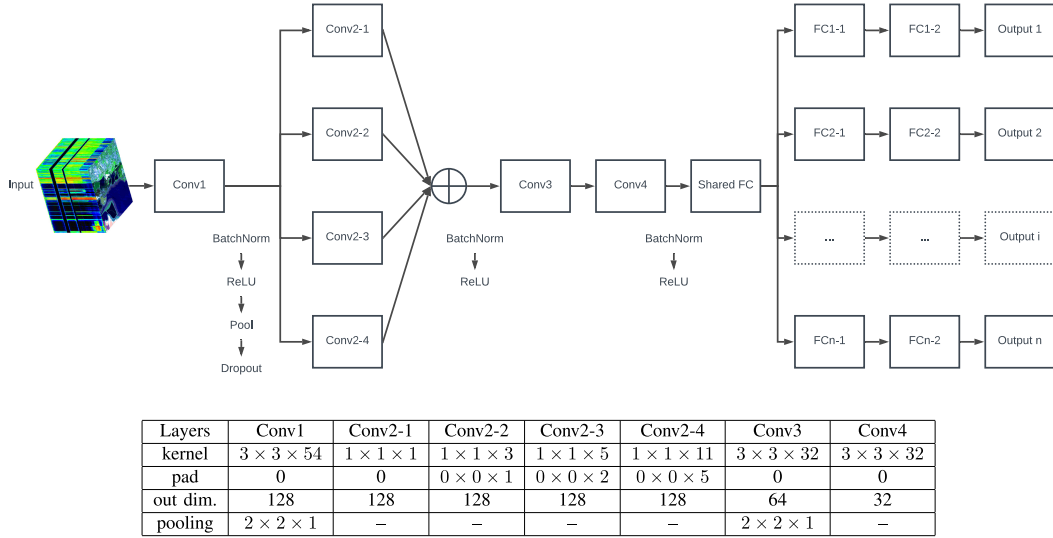


Fig. 4. Pipeline of the baseline TAIGA 3-D CNN model for hyperspectral forest variable prediction. The table summarizes the configurations of the Conv layers.

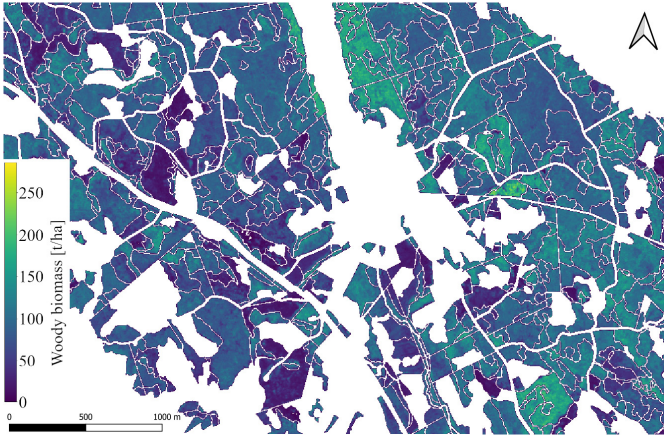


Fig. 5. Above-ground woody biomass prediction map for the forests in the area shown in Fig. 3.

The output from the Conv4 layer was flattened and fed to the shared FC layer, which produced a feature vector with a size of 512. Each task, i.e., one forest variable, had two task-specific stacked FC layers, which learned the mappings between the extracted features and the output for each individual task. The first FC layer took a 512-length feature vector from the shared FC layer and transformed it into a 200-length feature vector. If the task was a classification task with C classes, this vector was mapped to a vector of length C in the last FC layer and then fed into the Softmax function to compute the probabilities for all the classes. Otherwise, if the task was a regression task, the last FC layer was used to predict the output directly.

We used cross-entropy loss for the classification tasks and mean squared error loss for the regression tasks. The model was trained with an optimized batch size of 32 samples whose maximum value was determined by the memory constraints together with the image patch size. The learning rate was optimized in our experiments as described in [53], where the

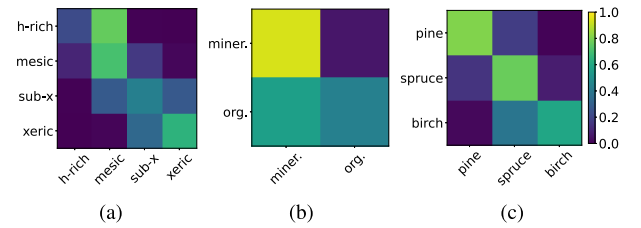


Fig. 6. Relative confusion matrices of the categorical forest variables. (a) Fertility class. (b) Soil class. (c) Main tree species. Overall and mean class accuracies are given in Table III. Abbreviations: h-rich = herb-rich, sub-x = sub-xeric, miner. = mineral, org. = organic.

value 10^{-4} was found to be optimal in the range $[10^{-7}, 10^{-1}]$. We used the Adam optimizer with the weight decay η at 10^{-4} , ρ_1 at 0.9 and ρ_2 at 0.999, which are commonly used as their default values.

Since the data are imbalanced, we used sampling with synthetic minority class samples as described in more detail in [53] to improve the classification performance of the categorical variables. For oversampling the minority classes, we used radiation-based and mixture-based methods as described in [20]. Class rectification loss [49] and focal loss [50] are the two loss functions that were used to mitigate the class-imbalance problem. We used the hyperparameter values recommended in the original papers for these loss functions. With the focal loss, our model achieved better results on our multiclass multilabel classification tasks (see Pham [53] for details). Oversampling seemed to lead to model overfitting by seeing rare instances too often.

The code for the model and the baseline experiments for the TAIGA benchmark are available in the *taiga-baseline* directory in <https://github.com/aalto-cbir/AIROBEST>.

IV. BASELINE RESULTS

Our 3-D CNN model produced predictions of the 13 categorical and continuous variables for each data pixel in test set

sample. As an example, a subset of the map for above-ground woody biomass is shown in Fig. 5. The stand-wise predictions, used for accuracy assessment, were obtained as arithmetic means of the pixel-wise values for continuous variables or as majority estimates for the categorical ones.

For the categorical variables, we computed the overall (micro) and the mean class (macro) accuracies and formed the confusion matrices (see Fig. 6). The micro accuracy tends to favor the model's good performance for the majority class or classes, whereas the macro accuracy gives equal weight also to the minority classes. For the continuous variables, we used the basic statistics commonly used in forest remote sensing studies, such as the root-mean-square error (RMSE), relative RMSE (rRMSE), relative mean bias (rBias), and coefficient of determination (R^2). Additionally, we computed the widths of the 90% and 95% confidence intervals for the continuous variables. Additional experiments and their results in the same setting can be found in [46].

The retrieval accuracy metrics presented in Table III can be considered very good for forest variable retrieval from EO data [55] with $R^2 \geq 0.7$. The accuracy was generally better than the results obtained in a previous study based on the same data [54], but excluding spatial information. The most accurately predicted variable was tree height (rRMSE 12%). Tree species composition was predicted with the largest uncertainty—the 95% confidence interval was approximately 50 percentage points for all three species. Besides species composition, the rRMSE value for stand density (41%) was higher than that of other structural variables.

For two selected key structural variables, woody biomass and mean tree height, we can see that the estimated values lie in several clusters along the diagonal of the predicted versus target plot (Fig. 7) and the tails of the error histograms fall off reasonably rapidly (Fig. 8).

The histograms of field data on the variables are not uniform, indicating that the management regime of forests has undergone some changes during the rotation period: the tree height distribution has minima at approximately 12 and 22 m (Fig. 7). The structure in the target tree height distribution is not fully represented in the estimated height with an underestimation of the tree heights for the tallest stands. A similar effect of smoothing of the parameter distribution is also visible for woody biomass.

The overall accuracies varied largely among the categorical variables. The largest values were generally on the main diagonal of the confusion matrices (Fig. 6) with the exception of mesic sites predominantly predicted to be herb-rich [Fig. 6(a)].

We also wanted to verify that the multitask learning approach is not only beneficial in the sense of efficiency in training a single network model instead of multiple ones, but also leads to prediction accuracies that are not much degraded. First, we studied the multitask predictions of the categorical and continuous variables only, so practically having two models instead of one. Second, we ran experiments where we predicted each categorical and each continuous variable separately, which is the basic setup of nonmultitask learning. When only the categorical variables were predicted, the average overall micro accuracy was 78.7% and the mean class

macro accuracy 75.3%. It thus seems that multitask learning of all variables together has been beneficial for the micro accuracy but not for macro accuracy. For the prediction of only all the continuous variables, the rRMSE value was 23.6% and the R^2 value 0.79. These are both slightly better than those in Table III, but the obtained average rBias value -3.1% is worse than the result in the table. Among the single categorical variables, we chose to evaluate predictions of the fertility class and obtained micro and macro accuracies of 63.0% and 52.9%, respectively, which can be seen to be inferior to the multitask prediction results. Correspondingly, basal area was predicted as a single-task manner and the rRMSE, rBias, and R^2 values were 14.2%, -4.0% , and 0.76, respectively, and again the rRMSE and R^2 values were somewhat better than the multitask results. It thus seems that the multitask learning approach neither improves nor degrades the accuracy of the predictions results considerably, and the observed gains and losses due to it are approximately equal in size.

V. DISCUSSION

The value of the zettabytes of EO data produced by the numerous remote sensing satellites can only be utilized by integrating it with other information sources [56]. This is especially true for developing and validating novel machine learning tools for practical, real-life applications. These actions require a large number of reliably labeled pixels accompanied by sufficient spatial extent in the spatially registered imagery to allow spatial separation of the training and testing subsets. Thankfully, many national policies answer this need by opening up geospatial datasets, such as the high-quality forest variable data forming a part of TAIGA used in this study.

The forestry dataset used in this study includes both continuous and categorical variables and is tightly related to the real-world problems faced by remote sensing scientists. The forest variables in Table I are relevant not only to the forestry community but also to biologists, Earth system scientists, and many others. A simple classification approach, using categorical variables, does not answer the information needs for many important mapping activities. The multitask learning approach presented here was chosen to address these gaps. With our experiments, we also showed that the multitask approach gave prediction accuracies comparable to that of single-task learning while providing the benefit of hugely reduced complexity in model training.

Furthermore, the amount of published studies on multitask learning in remote sensing is still fairly limited. We release the TAIGA dataset openly and propose to use it as a benchmark for hyperspectral multitask learning in remote sensing, hoping to encourage more activity and development in this field of research.

The results presented here are good compared with the general accuracy obtained for forests with remote sensing [55]. Utilizing both spatial and spectral information yielded better results compared with earlier Gaussian process regression-based results with the same datasets and evaluation procedure [54]. Use of structural information improved forest variable retrieval somewhat, in line with the analysis by

TABLE III

BASELINE STAND LEVEL RESULTS WITH A MULTITASK 3-D CNN MODEL: OVERALL AND MEAN CLASS ACCURACIES (ACC) FOR CATEGORICAL VARIABLES; RMSE, rRMSE, rBIAS, COEFFICIENT OF DETERMINATION (R^2), AND WIDTHS OF THE 90% AND 95% CONFIDENCE INTERVALS FOR CONTINUOUS VARIABLES. THE REFERENCE RESULTS IN THE RIGHT-MOST COLUMN ARE BY HALME *et al.* [54] USING THE SAME DATA

Variable #	Variable name	Accuracy metrics						Reference results [55]			
		overall ACC (micro)			mean class ACC (macro)			RMSE		R^2	
<i>Categorical</i>											
1	Fertility class	70.3 %			53.6 %						
2	Soil class	91.0 %			78.2 %						
3	Main tree species	87.4 %			83.7 %						
mean		82.9 %			71.8 %						
<i>Continuous</i>											
		RMSE	rRMSE	rBias	R^2	90%	95%	RMSE	rRMSE	rBias	R^2
4	Basal area [m ² /ha]	3.60	16%	1%	0.69	11.54	13.85	3.85	17%	0%	0.67
5	Mean DBH [cm]	2.69	12%	-1%	0.78	7.23	10.60	—	—	—	—
6	Stem density [1/ha]	380	41%	4%	0.73	695	1412	—	—	—	—
7	Mean height [m]	2.03	12%	1%	0.77	6.08	7.83	2.60	15%	-2%	0.66
8	Percentage of pine	12.54	25%	-1%	0.83	36.42	52.39	—	—	—	—
9	Percentage of spruce	11.60	32%	1%	0.77	34.99	48.93	—	—	—	—
10	Percentage of birch	6.77	55%	-1%	0.73	23.03	27.84	—	—	—	—
11	Woody biomass [t/ha]	19.92	20%	0%	0.76	64.65	81.91	21.78	28%	-3%	0.62
12	Leaf area index	1.02	21%	1%	0.79	3.20	3.83	—	—	—	—
13	Effective leaf area index	0.69	21%	1%	0.81	2.05	2.76	0.66	20%	0%	0.83
mean		—	25%	1%	0.77	—	—	—	—	—	—

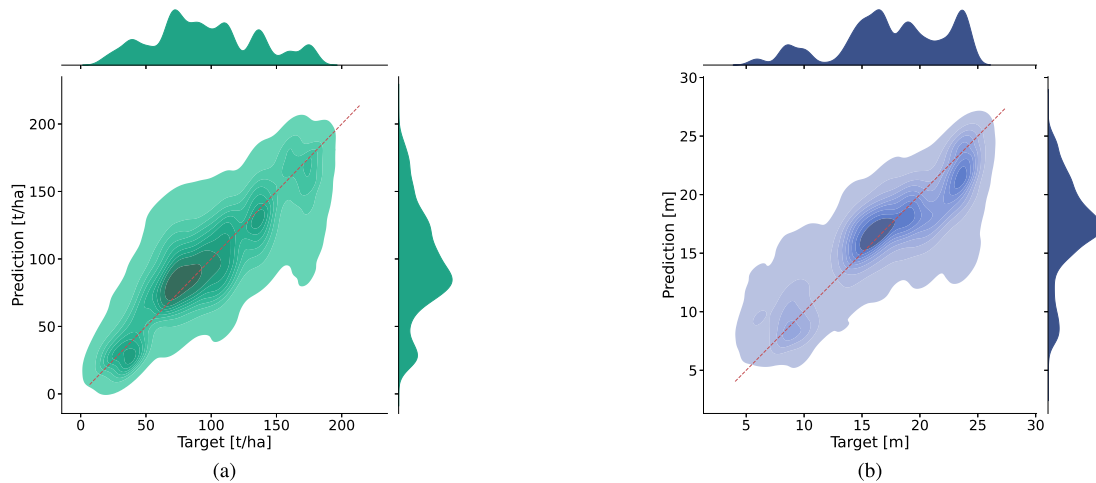


Fig. 7. Prediction accuracy of continuous forest variables. (a) Woody biomass. (b) Mean tree height.

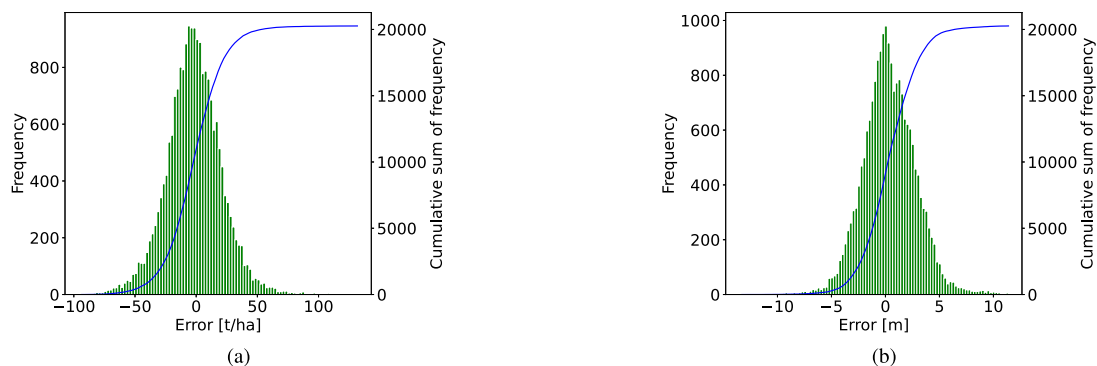


Fig. 8. Prediction error histogram for selected continuous forest variables. (a) Woody biomass, signed. (b) Mean tree height, signed.

White *et al.* [1]. High spatial and spectral resolution data will not be available jointly on a large scale, at least not in an operational fashion, in the foreseeable future. The relatively small advantage of high-resolution imagery does not likely

encourage the development of specialized systems for forest monitoring with VHR HSI instruments. However, operational HSI on a global scale is becoming a reality with high-resolution (pixel size 20–30 m) instruments [57]. While there

are parts of our planet for which VHR imagery is not readily available [58], the availability of these data is improving literally by the day. We suggest that combining VHR and HSI data from different sensors with advanced machine learning approaches may considerably improve the robustness of forestry data, especially for regions where national forest inventory data are not available.

The 3-D CNN model used to produce the baseline results of this article was the outcome of a series of experiments and improvements [53] that were left out from this article for the sake of brevity. The most important design choice in our CNN model was the use of multitask learning in one network instead of devising multiple separate models, one for each output forest variable. We are currently in the process of improving our CNN model for both the categorical and continuous variables. In particular, we look forward to studying the performance of the novel transformer architectures [59].

As with many natural observations, TAIGA data are heavily imbalanced, as is seen from the confusion matrices (Fig. 6). Furthermore, the reference forestry data have their limitations. It is based on a large number of forest plot measurements interpolated by mainly airborne laser scanning, which means that some inaccuracies are unavoidable. Unfortunately, only a very small number of the plots were within the HSI area, making their direct use impossible. Airborne HSI data are acquired even for a relatively large area within hours to days, depending on weather. Satellite sensors acquire their imagery near-instantaneously. Field plots, on the other hand, may take months to years to acquire on a regional scale, with national measurements spread over several years. The different temporal scales direct toward the use of well-maintained forest variable databases, which are updated based on forest management action notices and growth models, and can be used to retrieve forest information for the instance of image acquisition. The use of stand-level forest variable estimates and accuracy metrics in this study was motivated by the purpose of the FFC forestry data: they are computed to minimize errors at stand level, driven by the requirements of practical forest management. National forest inventories, for example, are designed to minimize bias at regional to national scales.

Naturally, a machine learning algorithm cannot exceed the (largely unspecified) error in the stand-level forest variables and the room for improving prediction accuracy is limited. We know that tree height is very reliably estimated from airborne laser scanning data in Finland with rRMSE values below 7% [60]. An rRMSE value of 12% for this variable (Table III) indicates that there is room for improvement. The indirect nature of the reference data also distorts the distributions of the forest variables. Still, a lot of structure is retained in the FFC data (Fig. 7), and some—although clearly less—reproduced by the baseline retrieval algorithm. In addition to the accuracy metrics in Table III, the structure in the predicted distributions is a prerequisite for a successful EO-based forest variable retrieval algorithm.

Up-to-date and accurate forest data are needed for almost all regions of the globe. Technical developments are taking place, which will make hyperspectral (CHIME [57]; EnMAP [61]; PRISMA [62]) and VHR EO data available everywhere,

promising improvements for future forest-monitoring systems. Machine learning can provide data-processing algorithms capable of handling the large volume of high-dimensional data. To develop and test these algorithms, datasets such as TAIGA presented here will be essential. Our results, although promising, need to be further improved and extended to be truly applicable on national to continental scales. To this end, it would be beneficial to release other benchmark datasets for multitask learning in different biomes. We hope that the TAIGA dataset and the baseline results presented in this article will support improvement of the tools required to implement them.

VI. CONCLUSION

We have presented the TAIGA dataset containing a hyperspectral image with more than 70 million pixels labeled with both categorical and continuous forest variables. The dataset is available through the Fairdata IDA service (<https://ida.fairdata.fi>). We have demonstrated a baseline system with a multitask convolutional neural network, which makes use of the spectral and spatial information in the data, available in the *taiga-baseline* directory in <https://github.com/aalto-cbir/AIROBEST>. The estimations presented in this article are more accurate than the ones made with only spectral information. The openly available data and the presented baseline results will help to make better use of the ongoing improvements in EO technology.

ACKNOWLEDGMENT

The authors would like to thank CSC—IT Center for Science, Finland, for computational resources.

REFERENCES

- [1] J. C. White, N. C. Coops, M. A. Wulder, M. Vastaranta, T. Hilker, and P. Tompalski, "Remote sensing technologies for enhancing forest inventories: A review," *Can. J. Remote Sens.*, vol. 42, no. 5, pp. 619–641, 2016, doi: [10.1080/07038992.2016.1207484](https://doi.org/10.1080/07038992.2016.1207484).
- [2] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jan. 2016.
- [3] J. E. Ball, D. T. Anderson, and C. S. Chan, "Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community," *J. Appl. Remote Sens.*, vol. 11, no. 4, pp. 11–54, 2017, doi: [10.1117/1.JRS.11.042609](https://doi.org/10.1117/1.JRS.11.042609).
- [4] X. X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.
- [5] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
- [6] D. Tuia, M. Volpi, M. D. Mura, A. Rakotomamonjy, and R. Flamary, "Automatic feature learning for spatio-spectral image classification with sparse SVM," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6062–6074, Oct. 2014.
- [7] Y. Gu, T. Liu, X. Jia, and J. A. Benediktsson, "Nonlinear multiple kernel learning with multiple-structure-element extended morphological profiles for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3235–3247, Jun. 2016.
- [8] Y. Gu, J. Chanussot, X. Jia, and J. A. Benediktsson, "Multiple Kernel learning for hyperspectral image classification: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6547–6565, Nov. 2017.
- [9] G. Fu, C. Liu, R. Zhou, T. Sun, and Q. Zhang, "Classification for high resolution remote sensing imagery using a fully convolutional network," *Remote Sens.*, vol. 9, no. 5, p. 498, May 2017.

- [10] Y. Chen, L. Zhu, P. Ghamisi, X. Jia, G. Li, and L. Tang, "Hyperspectral images classification with Gabor filtering and convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2355–2359, Dec. 2017.
- [11] S. Mei, J. Ji, J. Hou, X. Li, and Q. Du, "Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4520–4533, Aug. 2017.
- [12] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "A new deep convolutional neural network for fast hyperspectral image classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 120–147, Nov. 2018.
- [13] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.
- [14] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 279–317, Dec. 2019.
- [15] P. Ghamisi *et al.*, "New frontiers in spectral-spatial hyperspectral image classification: The latest advances based on mathematical morphology, Markov random fields, segmentation, sparse representation, and deep learning," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 3, pp. 10–43, Sep. 2018.
- [16] V. Slavković, S. Verstockt, W. De Neve, S. Van Hoesche, and R. Van De Walle, "Hyperspectral image classification with convolutional neural networks," in *Proc. ACM Multimedia Conf.*, Oct. 2015, pp. 1159–1162.
- [17] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *J. Sensors*, vol. 2015, pp. 1–12, Jan. 2015.
- [18] Y. Li, H. Zhang, and Q. Shen, "Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sens.*, vol. 9, no. 1, p. 67, Jan. 2017.
- [19] H. Lee and H. Kwon, "Contextual deep CNN based hyperspectral classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 3322–3325.
- [20] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.
- [21] W. Shao and S. Du, "Spectral-spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Oct. 2016.
- [22] L. Liu, M. Ji, and M. Buchroithner, "Transfer learning for soil spectroscopy based on convolutional neural networks and its application in soil clay content mapping using hyperspectral imagery," *Sensors*, vol. 18, no. 9, p. 3169, Sep. 2018.
- [23] J. Pyo *et al.*, "A convolutional neural network regression for quantifying cyanobacteria using hyperspectral imagery," *Remote Sens. Environ.*, vol. 233, Nov. 2019, Art. no. 111350.
- [24] J. M. Leiva-Murillo, L. Gomez-Chova, and G. Camps-Valls, "Multitask remote sensing data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 151–161, Jan. 2013.
- [25] S. Liu and Q. Shi, "Multitask deep learning with spectral knowledge for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 12, pp. 2110–2114, Dec. 2020.
- [26] S. Liu, Q. Shi, and L. Zhang, "Few-shot hyperspectral image classification with unknown classes using multitask deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5085–5102, Jun. 2021.
- [27] S. Jia, B. Deng, J. Zhu, X. Jia, and Q. Li, "Superpixel-based multitask learning framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2575–2588, May 2017.
- [28] J. Verrelst *et al.*, "Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties—A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 108, pp. 273–290, Oct. 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0924271615001422>
- [29] J. Liang, J. Zhou, Y. Qian, L. Wen, X. Bai, and Y. Gao, "On the sampling strategy for evaluation of spectral-spatial methods in hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 862–880, Feb. 2017.
- [30] R. Hänsch, A. Ley, and O. Hellwich, "Correct and still wrong: The relationship between sampling strategies and the estimation of the generalization error," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 3672–3675.
- [31] J. Lange, G. Cavallaro, M. Götz, E. Erlingsson, and M. Riedel, "The influence of sampling methods on pixel-wise hyperspectral image classification with 3D convolutional neural networks," in *Proc. IGARSS*, Jul. 2018, pp. 2087–2090.
- [32] S. Li, Q. Hao, G. Gao, and X. Kang, "The effect of ground truth on performance evaluation of hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 7195–7206, Dec. 2018.
- [33] C. Debes *et al.*, "Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2405–2418, Jun. 2014.
- [34] Y. Xu *et al.*, "Advanced multi-sensor optical remote sensing for urban land use and land cover classification: Outcome of the 2018 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1709–1724, Jun. 2019.
- [35] N. Yokoya and A. Iwasaki, "Airborne hyperspectral data over Chikusei," Space Appl. Lab., Univ. Tokyo, Japan, Tech. Rep. SAL-2016-05-27, May 2016.
- [36] P. Pirinen, H. Simola, J. Aalto, J. Kaukoranta, P. Karlsson, and R. Ruuhela, *Tilastoja Suomen Ilmastosta 1981–2010* (Reports). Helsinki, Finland: Finnish Meteorol. Inst., 2012, vol. 1.
- [37] V. Markiet, R. Hernández-Clemente, and M. Möttöus, "Spectral similarity and PRI variations for a boreal forest stand using multi-angular airborne imagery," *Remote Sens.*, vol. 9, no. 10, p. 1005, Sep. 2017. [Online]. Available: <http://www.mdpi.com/2072-4292/9/10/1005>
- [38] M. Maltamo, P. Packalén, E. Kallio, J. Kangas, J. Uuttera, and J. Heikkilä, "Airborne laser scanning based stand level management inventory in Finland," in *Proc. SilviLaser, 11th Int. Conf. LiDAR Appl. Assessing Forest Ecosyst., Univ. Tasmania, Aust. Bonn, Germany: Conf. Secretariat*, Oct. 2011, pp. 1–10. [Online]. Available: <http://www.iufro.org/download/file/8239/5065/40205-silvilaser2011.pdf>
- [39] M. Holopainen *et al.*, "Metsätalouteen, maatalouteen ja vesistöihin liittyviä paikka-tietosovelluksia," in *Geoinformatiikka Luonnonvarojen Hallinnassa*. Helsinki, Finland: Univ. Helsinki, 2015, no. 7, ch. 14, pp. 121–130. [Online]. Available: <http://hdl.handle.net/10138/166765>
- [40] K. Kuusela, "The 5th national forest inventory in Finland," *Communications Instituti Forestalis Fenniae*, vol. 69, no. 4, pp. 5–72, 1969.
- [41] A. Raunio, A. Schulman, and T. Kontula, Eds., "Suomen luontotyyppien uhanalaisuus—Osa 2: Luontotyyppien kuvaukset. [Assessment of the threatened habitat types in Finland—Part 2: Habitat type descriptions]," Suomen ympäristökeskus, Helsinki, Suomen ympäristö 8/2008, p. 574.
- [42] J. Repola, "Biomass equations for birch in Finland," *Silva Fennica*, vol. 42, no. 4, pp. 605–624, 2008. [Online]. Available: <https://www.silvafennica.fi/article/236>
- [43] J. Repola, "Biomass equations for scots pine and Norway spruce in Finland," *Silva Fennica*, vol. 43, no. 4, pp. 625–647, 2009. [Online]. Available: <https://www.silvafennica.fi/article/184>
- [44] T. Majasalmi, M. Rautiainen, P. Stenberg, and P. Lukeš, "An assessment of ground reference methods for estimating LAI of boreal forests," *Forest Ecol. Manage.*, vol. 292, pp. 10–18, Mar. 2013.
- [45] W. Rouse, H. Haas, J. A. Schell, and W. Deering, "Monitoring vegetation systems in the Great Plains with ERTS," in *Proc. 3rd Earth Resour. Technol. Satell.-1 Symp., Tech. Presentations*, vol. 1, S. C. Freden, E. P. Mercanti, and M. Becker, Eds. Washington, DC, USA: NASA, 1973, pp. 309–317.
- [46] T. Cu, "Prediction of forest variables and its reliability with a hyperspectral deep learning model," M.S. thesis, Aalto Univ., School Sci., Espoo, Finland, 2021. [Online]. Available: <http://urn.fi/URN:NBN:fi:aalto-202108298608>
- [47] M. Mottus, M. Molinier, E. Halme, H. Cu, and J. Laaksonen, "Patch size selection for analysis of sub-meter resolution hyperspectral imagery of forests," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Brussels, Belgium, Jul. 2021.
- [48] M. He, B. Li, and H. Chen, "Multi-scale 3D deep convolutional neural network for hyperspectral image classification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3904–3908.
- [49] Q. Dong, S. Gong, and X. Zhu, "Class rectification hard mining for imbalanced deep learning," 2017, *arXiv:1712.03162*.
- [50] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," 2017, *arXiv:1708.02002*.
- [51] Z. Chen, V. Badrinarayanan, C.-Y. Lee, and A. Rabinovich, "GradNorm: Gradient normalization for adaptive loss balancing in deep multitask networks," 2017, *arXiv:1711.02257*.
- [52] A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," 2017, *arXiv:1705.07115*.

- [53] P. Pham, “Deep learning methods for modelling forest biomass and structures from hyperspectral imagery,” M.S. thesis, Aalto Univ., School Sci., Espoo, Finland, 2019. [Online]. Available: <http://urn.fi/URN:NBN:fi:aalto-201906234142>
- [54] E. Halme, P. Pellikka, and M. Mottus, “Utility of hyperspectral compared to multispectral remote sensing data in estimating forest biomass and structure variables in Finnish boreal forest,” *Int. J. Appl. Earth Observ. Geoinf.*, vol. 83, Nov. 2019, Art. no. 101942. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0303243419304179>
- [55] H. Astola, T. Häme, L. Sirro, M. Molinier, and J. Kilpi, “Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region,” *Remote Sens. Environ.*, vol. 223, pp. 257–273, Mar. 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425719300252>
- [56] C. Vaduva, M. Iapaolo, and M. Datcu, *A Scientific Perspective on Big Data in Earth Observation*. Cham, Switzerland: Springer, 2020, pp. 155–188, doi: [10.1007/978-3-030-43981-1_8](https://doi.org/10.1007/978-3-030-43981-1_8).
- [57] J. Nieke and M. Rast, “Towards the copernicus hyperspectral imaging mission for the environment (CHIME),” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 157–159.
- [58] M. Lesiv *et al.*, “Characterizing the spatial and temporal availability of very high resolution satellite imagery in Google Earth and Microsoft Bing maps as a source of reference data,” *Land*, vol. 7, no. 4, p. 118, Oct. 2018. [Online]. Available: <https://www.mdpi.com/2073-445X/7/4/118>
- [59] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, “Transformers in vision: A survey,” 2021, *arXiv:2101.01169*.
- [60] E. Kotivuori, L. Korhonen, and P. Packalen, “Nationwide airborne laser scanning based models for volume, biomass and dominant height in Finland,” *Silva Fennica*, vol. 50, no. 4, p. 1567, 2016. [Online]. Available: <https://www.silvafennica.fi/article/1567>
- [61] L. Guanter *et al.*, “The EnMAP spaceborne imaging spectroscopy mission for earth observation,” *Sensor*, vol. 7, no. 7, pp. 8830–8857, Jul. 2015. [Online]. Available: <https://www.mdpi.com/2072-4292/7/7/8830>
- [62] L. Candela, R. Formaro, R. Guarini, R. Loizzo, F. Longo, and G. Varacalli, “The PRISMA mission,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 253–256.



Matti Möttus received the M.Sc. degree in physics and the Ph.D. degree in environmental physics from the University of Tartu, Tartu, Estonia, in 2000 and 2004, respectively.

He has worked with the Tartu Observatory, Tõravere, Estonia, and the University of Helsinki, Helsinki, Finland. He was an Academy of Finland Research Fellow with the Department of Geosciences and Geography, University of Helsinki, and VTT Technical Research Centre of Finland, Espoo, Finland, from 2013 to 2018. He is currently a

Principal Scientist with the VTT Technical Research Centre of Finland, where he is interested in imaging spectroscopy, mapping of forest structure and function, and radiative transfer in vegetation canopies.



Phu Pham received the M.Sc. degree in Big Data and large-scale computing and a minor in machine learning and data mining from Aalto University, Espoo, Finland, in 2019. He is currently pursuing the Ph.D. degree with the Department of Computer Science, Purdue University, West Lafayette, IN, USA.

His research interests include computer vision, natural language processing, and robotics.



Elis Halme received the master’s degree in geoinformatics (major) from Aalto University, Espoo, Finland, in 2018, where he is currently pursuing the Ph.D. degree.

He is currently a Research Scientist with the VTT Technical Research Centre of Finland, Espoo, working with Earth observation, hyperspectral data, and machine learning to retrieve information on boreal forest.



Matthieu Molinier (Member, IEEE) received the engineering degree from École Nationale Supérieure de Physique de Strasbourg (ENSPS), Illkirch-Graffenstaden, France, and the M.Sc. degree in image processing from the Université Louis Pasteur (ULP), Strasbourg, France, both in 2004.

Since then, he has been working with VTT, Espoo, Finland, as a Research Scientist in Earth observation, with a main focus on machine learning and change detection for optical satellite images. He is the author of 60 articles in scientific journals and conferences.

His research interests include deep learning for multispectral and hyperspectral images, unsupervised change detection, and satellite image time-series analysis, with applications to environment monitoring.



Hai Cu is currently pursuing the master’s degree (major in machine learning, data science, and artificial intelligence) with Aalto University, Espoo, Finland, with the thesis topic focusing on the reliability of the forest variable predictions from deep learning hyperspectral model.

He has participated in AIROBEST Project since early 2020.



Jorma Laaksonen (Senior Member, IEEE) received the Doctor of Science in Technology degree from the Helsinki University of Technology, Helsinki, Finland, in 1997.

He is currently a Senior University Lecturer with the Department of Computer Science, Aalto University School of Science, Espoo, Finland. He is the author of 30 journals and 160 conference articles on pattern recognition, statistical classification, machine learning, and neural networks. His research interests are in content-based multimodal information analysis and retrieval and computer vision.

Dr. Laaksonen is an Associate Editor of *Pattern Recognition Letters* and a Founding Member of the SOM and LVQ Programming Teams and the PicSOM Development Group.